PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-293860
(43)Date of publication of application : 04.11.1998

(51)Int.CL G06T 13/00
G10L 3/00

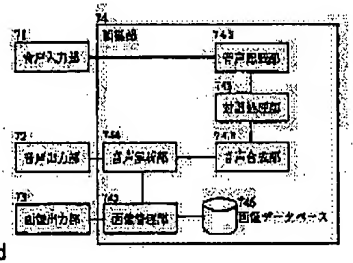
(21)Application number : 09-216377
(22)Date of filing : 11.08.1997

(71)Applicant : NIPPON TELEGR & TELEPH CORP <NTT>
(72)Inventor : MATSUURA MICHIAKI
KURA TSUNEKO
OSHIMA TAKASHI
WATANABE NOBUYUKI
KANAYAMA HIDEAKI

(30)Priority
Priority number : 09 38760 Priority date : 24.02.1997 Priority country : JP

(54) PERSON IMAGE DISPLAY METHOD AND DEVICE USING VOICE DRIVE

(57)Abstract:
PROBLEM TO BE SOLVED: To artificially attain an almost natural conversation state with no use of any special device, etc., by changing at least one of shakes of a mouth, head, etc., of a person face image based on the loudness of human voices which are obtained in sequence and at each fixed time interval.
SOLUTION: A voice analysis part 744 picks up the changes of voices received from a voice synthesizing part 743 at each prescribed time interval and calculates the mean value of loudness of voices, etc. Then the time series information on each of mean value of voice loudness divided at each time interval of voices is outputted to an image management part 745. The part 745 successively selects the prescribed proper images out of an image data base in response to each information and outputs these images to an image output part 73 to successively display them. These images primarily show an answering person to the voices of a speaker, and the mouths, attitudes, etc., of one or more persons are operated in accordance with the voices produced from a voice output part 72.



(19)日本国特許庁(JP)　(12)公開特許公報（A）　(11)特許出願公開番号
特開平10－293860
(43)公開日　平成10年(1998)11月4日

(51)Int. Cl.[6]　識別記号　F I
G06T 13/00　G06F 15/62 340 D
G10L 3/00　G10L 3/00 S

審査請求　未請求　請求項の数11　OL　(全14頁)

(21)出願番号　特願平9-216377
(22)出願日　平成9年(1997)8月11日
(31)優先権主張番号　特願平9-38760
(32)優先日　平9(1997)2月24日
(33)優先権主張国　日本（JP）

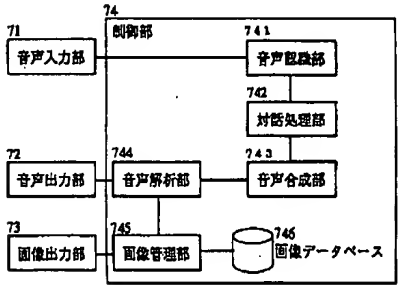
(71)出願人　000004226
日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号
(72)発明者　松浦　道明
東京都新宿区西新宿三丁目19番2号　日本電信電話株式会社内
(72)発明者　倉　恒子
東京都新宿区西新宿三丁目19番2号　日本電信電話株式会社内
(72)発明者　大島　孝
東京都新宿区西新宿三丁目19番2号　日本電信電話株式会社内
(74)代理人　弁理士　秋田　収喜

最終頁に続く

(54)【発明の名称】音声駆動を用いた人物画像表示方法およびその装置

(57)【要約】
【課題】　専門の技術あるいは装置を必要としなくてもほぼ自然的に対話をしている状態を疑似的に実現できる。
【解決手段】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、前記人物の発する音声の一定時間間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像における口、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させる。

図1



1

【特許請求の範囲】

【請求項１】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、

前記人物の発する音声の一定時間間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像におけるロ、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする音声駆動を用いた人物画像表示方法。

【請求項２】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示装置において、

前記人物の顔画像におけるロ、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つが変化された各画像を格納する格納手段と、

前記人物の発する音声の一定時間間隔ごとの各音声の大きさを順次算出する算出手段と、

この算出手段によって算出された算出値に応じて、前記格納手段に格納された各画像を順次選定する画像選定手段と、

この画像選定手段によって選定された各画像を順次表示する表示手段と、を備えることを特徴とする音声駆動を用いた人物画像表示装置。

【請求項３】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、

前記人物の発する音声の一定時間間隔ごとに順次得られる各音声の差分に基づいて、該人物の顔画像におけるロ、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とする音声駆動を用いた人物画像表示方法。

【請求項４】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示装置において、

前記人物の顔画像におけるロ、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つが変化された各画像を格納する格納手段と、

前記人物の発する音声の一定時間間隔ごとの各音声の差分を順次算出する算出手段と、

この算出手段によって算出された差分値に応じて、前記格納手段に格納された各画像を順次選定する画像選定手段と、

この画像選定手段によって選定された各画像を順次表示する表示手段と、を備えることを特徴とする音声駆動を用いた人物画像表示装置。

【請求項５】　コンピュータに表示されている人物の発

2

する音声は、ネットワークを介して送られる音声であることを特徴とする請求項１および３のうちいずれか記載の人物画像表示方法。

【請求項６】　コンピュータに表示されている人物の発する音声は、ネットワークを介して送られる音声であることを特徴とする請求項２および４のうちいずれか記載の人物画像表示装置。

【請求項７】　コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、

前記人物の発する音声から一定時間間隔毎に得られる有音／無音の判定に応じて、ロの大きさ、目の瞬き、頭の揺れ、うなずきの少なくとも一つを制御することを特徴とする音声駆動を用いた人物画像表示方法。

【請求項８】　有音判定の際に、予め定められた周期あるいはランダムな時間間隔でロの大きさを変化させ、無音判定の際に、ロを閉じることを特徴とする請求項７記載の音声駆動を用いた人物画像表示方法。

【請求項９】　有音判定の際に、予め定められた周期あるいはランダムな時間間隔で目を１回瞬かせ、無音判定の際に、予め定められた周期あるいはランダムな時間間隔で目を複数回瞬かせることを特徴とする請求項７記載の音声駆動を用いた人物画像表示方法。

【請求項１０】　有音判定の際に、予め定められた周期あるいはランダムな時間間隔で頭を動かせることを特徴とする請求項７記載の音声駆動を用いた人物画像表示方法。

【請求項１１】　無音判定の際に、予め定められた周期あるいはランダムな時間間隔で縦方向に頭を動かせるか、首を傾けるか、首を横に振らせることを特徴とする請求項７記載の音声駆動を用いた人物画像表示方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は音声駆動を用いた人物画像表示方法およびその装置に係り、たとえばコンピュータとインタラクティブに対話を行うシステムにおいて、音声の情報を用いて人物の画像の顔部分を変形させることにより、対話をしている状態を疑似的に実現する方法および装置に関する。

【０００２】

【従来の技術】この種の方法および装置としては、たとえばニューラルネットワークを用いて口の形を決定する方法が知られている（文献：宮下、佐藤、坂口、森島「仮想人物との対話を実現するための音声から画像への実時間メディア変換システムの研究」信学技法MVE-95-62,1996-03）。

【０００３】すなわち、音声の特徴パラメータである発話者の声道特性と放射特性を表しているLPCケプストラム係数から口形情報への変換をニューラルネットワーク

3

を用いて行い、これにより実際の口の形に近い表示を可能としている。

【０００４】

【発明が解決しようとする課題】しかしながら、このようにニューラルネットワークを用いた口形の決定に要する処理は、高度の専門の技術を要し、さらに、高度の専用の装置を必要とするという問題が指摘されるに到っている。

【０００５】そのため、このような問題を検討した結果、たとえばアニメーションや吹き替えの映画等を鑑賞する場合に明らかなように、必ずしも口形と実際の音声とが一致していなくても、話し始めで口が開き話し終わりで口が閉じることにより、ほとんどの観客が不自然さを感じない事実を考慮すると、必ずしも専門の技術あるいは装置を必要としなくてもほぼ自然的に対話をしている状態を疑似的に実現できることが判明した。

【０００６】本発明は、このような事情に基づいてなされたものであり、その目的は、専門の技術あるいは装置を必要としなくてもほぼ自然的に対話をしている状態を疑似的に実現できる音声駆動を用いた人物画像表示方法およびその装置を提供することにある。

【０００７】

【課題を解決するための手段】本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記の通りである。

【０００８】すなわち、コンピュータに表示されている人物の画像の少なくとも顔の部分がその人物の発する音声によって変化する音声駆動を用いた人物画像表示方法において、前記人物の発する音声の一定時間間隔ごとに順次得られる各音声の大きさに基づいて、該人物の顔画像におけるロ、頭の揺れあるいはうなずき、目のまばたきの少なくとも一つを変化させることを特徴とするものである。

【０００９】このように構成した音声駆動を用いた人物画像表示方法によれば、人物の表情において、たとえばその口の場合に、話し始めで口が開き話し、さらに終わりでロが閉じるようになり、また声の大きさに分類した数に応じてそれらの間で口の開きの大きさを変化させることができることから、ほとんどの対話者が不自然さを感じることなく観察することができるようになる。

【００１０】したがって、特に、複雑な手法を用いる必要がないことから、専門の技術あるいは装置を必要とすることなくほぼ自然的に対話をしている状態を疑似的に実現できるようになる。

【００１１】

【発明の実施の形態】以下、本発明による音声駆動を用いた人物画像表示方法および装置の実施例について図面を用いて説明する。

【００１２】実施例１．図１は、本発明による音声駆動を用いた人物画像表示装置の一実施例を示すブロック図

4

である。

【００１３】同図において、コンピュータとインタラクティブに対話を行う対話者の音声を入力する音声入力部７１があり、この音声入力部７１からの信号はコンピュータ側の制御部７４に取り込まれるようになっている。

【００１４】制御部７４では、まず、前記音声入力部７１からの信号を音声認識部７４１によって、その音声を認識し、かつその言葉の意味を判断するようになっている。

【００１５】そして、対話処理部７４２では対話者の意図を汲み取り、かつ、それに対する返答を判断するようになっている。

【００１６】音声合成部７４３では、前記対話処理部７４２によって判断された返答に相当する音声を作成するようになっている。この音声は強さの変化のない単調なものではなく、あたかも実際の感情を有する人間が話すようにその言葉の意味に応じて強弱（アクセントおよびイントネーション）が付けられたものとなっている。

【００１７】そして、この音声合成部７４３からの音声は音声解析部７４４を介して音声出力部７２に出力され、この音声出力部７２から前記返答に相当する音声が発せられるようになっている。

【００１８】前記音声解析部７４４では、音声合成部７４３からの音声に対してたとえばＴｉの時間間隔ごとに該音声の変化を取りだし、その強さの平均値の算出、あるいはその他の必要な演算処理を行うようになっている。

【００１９】そして、この音声解析部７４４で得られた音声の時間毎に区切られた各平均値等の時系列的な各情報は画像管理部７４５に出力され、この画像管理部７４５では、前記それぞれの情報に応じて画像データベース７４６から予め定められた適当な画像が順次選択され、この画像は画像出力部７３に出力されて順次表示されるようになっている。

【００２０】この画像は、前記対話者の音声に対して返答をする人物の画像（一人の場合もあるし、複数の場合もある）が主となり、その人物の口あるいはそぶり等は、前記音声出力部７２から発せられる音声に追従して（同じタイミングで）動くようになっている。

【００２１】次に、前記音声解析部７４４、画像管理部７４５、および画像データベース７４６の各機能を、基本アルゴリズムおよびそれをさらに拡張させた拡張アルゴリズムに基づいて以下説明する。

【００２２】基本アルゴリズムとして、まず、説明の便宜上口のみの大きさの異なる複数の画像を予め用意する。この画像は画像データベース７４６に格納されるものとなっている。

【００２３】そして、音声合成部７４３からの音声信号を音声解析部７４４によって、図２に示すように、ある一定時間の間隔Ｔｉごとに音声の大きさの平均値Ｓｎを

求める。

【0024】さらに、音声の大きさの各平均値Saに基づいて、図3に示すように、前記各画像の遷移先を決定する。

【0025】図3においては、4種類の口の大きさを持った画像を例示しており、たとえば音声入力がない場合にはすべて口を閉じた画像に遷移する。音声の大きさの最大値を4つに分け、その各々の区間に口の大きさの異なる画像を配置する。Saの値が変わることにより、口の大きさも変化する。このとき、口の大きさは急激に変わる場合もある。

【0026】そして、拡張アルゴリズムとして、前記基本アルゴリズムに無意識の動きである揺らぎを加えたものとなっている。

【0027】すなわち、その一部を図4に示すように、まず、基準となる顔の変化をB、頭を揺らした顔の変化をH、目を閉じた顔の変化をEとする。

【0028】そして、音声の大きさの平均Saが与えられた際には、次式（1）に示す確率に従ってあるパターンから別のパターンへ状態を変位させるようになっている。

【0029】

【数1】

$$P_{B\to B} + P_{B\to H} + P_{B\to E} = 1 \quad \cdots\cdots (1)$$

$P_{B\to B}$：基準となる顔から基準となる顔へ遷移する確率

$P_{B\to H}$：基準となる顔から頭を揺らした顔へ遷移する確率

$P_{B\to E}$：基準となる顔から目を閉じた顔へ遷移する確率

【0030】なお、この場合、上式（1）の各確率はその等式が満足できる限りどのような割合にしてもよいことはいうまでもない。

【0031】ここで、前記基本アルゴリズムについてさらに詳述すると、上述したように、ある一定時間間隔Tiごとに音声の平均値Saを求め、さらに、図5に示すように、ある時間T_nとT_{n+1}（$=T_n+T_s$）との平均値の差分をD（$=Sa_{n+1}-Sa_n$）とおく。

【0032】そして、この差分Dの大きさにより、図6に示すように、画像の遷移先を決定する。同図では、差分Dがたとえばあるしきい値ε内に収まる場合には自分自身に遷移するようにしている。差分Dがしきい値εを越え、かつその値が正の場合には口をさらに大きく開け※た画像に遷移し、負の場合には口をさらに小さく開けた画像に遷移するようにしている。

【0033】このようにすることによって、口の開閉は滑らかな動きができるようになる。

【0034】また、このような差分Dに基づく前記拡張アルゴリズムについてさらに詳述すると、図7に示すように、上述した拡張アルゴリズムの第n状態において、差分Dがしきい値εを越え、かつその値が正の場合には第（n＋1）状態への遷移は次式（2）に示す等式を満足して行われるようになっている。

【0035】

【数2】

$$P_{Bn\to Bn+1} + P_{Bn\to Hn+1} + P_{Bn\to En+1} = 1 \quad \cdots\cdots (2)$$

$P_{Bn\to Bn+1}$：基準の顔で口を開いた顔へ遷移する確率

$P_{Bn\to Hn+1}$：頭を揺らし口を開いた顔へ遷移する確率

$P_{Bn\to En+1}$：目を閉じて口を開いた顔へ遷移する確率

【0036】そして、差分Dが負の場合には第（n－1）状態への遷移は次式（3）に示す等式を満足して行われるようになっている。★【0037】

【数3】

$$P_{Bn\to Bn-1} + P_{Bn\to Hn-1} + P_{Bn\to En-1} = 1 \quad \cdots\cdots (3)$$

$P_{Bn\to Bn-1}$：基準の顔で口を閉じた顔へ遷移する確率

$P_{Bn\to Hn-1}$：頭を揺らし口を閉じた顔へ遷移する確率

$P_{Bn\to En-1}$：目を閉じて口を閉じた顔へ遷移する確率

【0038】差分Dがしきい値ε内にある場合には自分自身へ遷移し、また、音声の大きさが0の場合には第1状態に遷移することはいうまでもない。

【0039】このようにすることによって、しきい値εを越えた音声の変化に応じて、人物の顔はすべて別の顔の状態に遷移することとなり、話をしている時の動きを滑らかに表現させることができるようになる。

【0040】ここで、拡張アルゴリズムを使って顔画像を表す実験を行った結果を以下に説明する。

【0041】この実験における基本アルゴリズムとして、図8（a）に示すように、声の大きさの変化にともなって口の大きさが4段階に変化する顔画像を用意した。

【0042】そして、拡張アルゴリズムとして、図8（b）に示すように、上述した3パターンに、目を半分閉じたパターンを追加し、遷移先は一つ前の状態には戻らない設定とした。また、音声を受信している際には、瞬きをしたり、うなずく等の動作をランダムに表示するようにした。＊

＊【0043】そして、被験者10人に対して、それぞれ基本アルゴリズムに基づく画像と拡張アルゴリズムに基づく画像とを提示し、それにより得られた結果を評定尺度法（大串、中山、福田、”画質と音質の評価技術”昭晃堂：1991）に基づいて評価してみた。

【0044】すなわち、カテゴリー数を5とし、端極尺度値により各々の刺激の尺度値を求め、その結果は表1に示されるようになった。

【0045】

【表1】

表示アルゴリズム評価結果

Table 2 The Results of Algorithms

カテゴリー	単極尺度値	基本アルゴリズム	拡張アルゴリズム
非常によい	4	1	1
よい	3	1	3
普通	2	2	4
悪い	1	5	2
非常に悪い	0	1	0
刺激の両極尺度値		1.60	2.30

【0046】この表1から明らかなように、拡張アルゴリズムによる評価（尺度値＝2．3）が基本アルゴリズムの評価（尺度値＝1．6）より高いことが判る。

【0047】この理由は、基本アルゴリズムでは、顔の他の部分を動かさなかったため、口の動きだけが目だってしまい、時間を経るに従ってその不自然さが増大する状態に陥り易いからである。拡張アルゴリズムではこの点の不都合を充分に解決したものとなっている。

【0048】また、表示されている人物が複数人おり、そのうちの一人の話を聞いている人物の場合において、その口を閉じた状態の画像で、目を閉じた顔と目を開いた顔との表示をランダムに切り替えることによって、その人物の目の瞬きを実現できるようになっている。

【0049】このように構成した人物画像表示方法および装置の利用法としては、たとえば商品紹介、タウンガイド、あるいは交通情報表示等の情報提供が挙げられる。

【0050】たとえば交通情報の場合は、具体的に、交通手段（鉄道／バス／道路／飛行機…）、混み具合（所要時間／渋滞距離／渋滞原因…）、その他の情報（天気予報／工事情報／事故…）等である。

【0051】利用者は、たとえば、提示されているメニューから必要な項目を音声で入力するようになる。入力された音声は音声認識部で認識され、認識された言葉に対応する返答を合成／解析し、その結果に基づき画面に表示する顔を動かす。

【0052】音声合成部は、返答文に利用者や時間により変化する情報を組み込むために必要となるものであり、音声解析部では合成された音声からその大きさの変化を抽出する。

【0053】これらのサービスのやり取りはある程度定型化できるものであり、入力される音声に対する応答をあらかじめ想定することが可能となっている。

【0054】たとえば交通情報の例では、順調に動いている場合はにこやかな顔を選択し、「○時○○分現在、××は順調に動いております」といった内容を音声とともにに口や顔を動かす。

【0055】認識できなかった場合には再度入力を促す。この場合、恐縮している顔を選択し、「申し訳ありませんが、もう一度入力していただけますか」などといった音声とともにに口や顔を動かす。

【0056】すなわち、提示する言葉の内容が予め判っているため、その内容に合わせて表情を持つ顔を対応づけ、アルゴリズムにしたがってロや顔を動かすことにより、より自然に話しをしている状況を表現することができるようになる。

【0057】実施例2．図9および図10は、本発明のさらに拡張された他の実施例を説明する説明図である。

【0058】ここで、実施例の説明をする前に、対話の際における対話者の自然な動作についてさらに考察してみる。

【0059】まず、対話というものは、一方的に言葉を

9

相手に投げるものでなく、相手の反応を見ながら行うものである。

【0060】対話がスムーズになされている状態を表現するためには、話をしている側の顔の部分にそれなりの動きをもたせることはもちろんのこと、聞いている側の顔の部分もそれなりに表現させる必要がある。

【0061】そこで、対話の際の話をしている側および聞いている側のそれぞれの顔の部分の動きを観察すると次のような動作傾向があることが判る。

【0062】（1）話をしている側の動作
・頭が動くことが多い。
・瞬きの回数は一回が多い。
・瞬きから瞬きの間隔が比較的長い。

【0063】（2）話を聞いている側の動作
・瞬きは複数回連続して行われることが多い。
・相槌をうつため、縦方向に頭が動く。
・疑問を感じた際に、首を傾ける。
・納得できない時に首を横に振る。

【0064】図9および図10に示す実施例では、対話者のこのような動作傾向を出力制御に反映させていることにある。

【0065】すなわち、コンピュータに表示されている人物が、話をしているのか（有音）、あるいは話を聞いているのか（無音）の判定を行うとともに、それぞれの場合における顔の部分の動作を、口の大きさ、目の瞬きに着目して制御するようにしている。

【0066】図9、図10では、図面の横方向に時間経過を示し、上段から順次、①有音／無音の判定タイミング、②有音／無音出力、③口画像出力、④目画像出力、⑤顔画像出力を示している。そして、図9は、主として有音判定時を示し、図10は、主として無音判定時を示している。なお、図9、図10は、それらが合わさって一つの図を構成するようになっている。

【0067】ここで、①有音／無音の判定タイミングでは、有音／無音の判定が一定周期T_Jでなされている。

【0068】②有音／無音出力では、有音判定時に、入力音声が無音レベルから有音レベルに変化したことにより、無音から有音に変化することを示し、また無音判定時に、入力音声が有音レベルから無音レベルに変化したことにより、有音から無音に変化することを示している。

【0069】有音／無音の判定は、たとえば判定間隔T_Jの中に存在する標本化音声の平均電力を計算し、その電力が予め定めたしきい値を越えるかどうかで行うようにしている。

【0070】③口画像出力では、有音判定時に、有音が開始してから一定周期（T_M）でロの大きさが変化することを示し、また無音判定時に、直ちにロが閉じることを示している。

【0071】④目画像出力では、有音判定時に、有音が

10

開始してからランダムに瞬きが発生し、その後は一定の周期（T_M）で瞬きが繰り返すことを示し、また無音判定時に、無音が開始してからランダムに瞬きが発生し、その後は一定の周期（T_E'）で瞬きが繰り返すことを示している。

【0072】この場合、無音では、瞬きの回数が2回連続して行われることを示している。

【0073】⑤顔画像出力、口画像と目画像を合成した結果を示している。

【0074】なお、上述したタイムチャートでは、有音から無音を検出した段階で、直ちに口を閉じるようにして示したが、このようにすることに限定されることはなく、検出時から時間をかけて漸次に口を閉じる制御を行うようにしてもよいことはいうまでもない。

【0075】また、口画像出力において、口の開閉の繰返し周期T_Mを一定として説明したものであるが、このようにすることに限定されることはなく、ランダムの時間幅で口の開閉を繰り返すようにしてもよいことはいうまでもない。

【0076】さらに、目画像出力において、目の瞬きの周期T_E, T_E'を一定として説明したものであるが、このようにすることに限定されることはなく、ランダムの周期で目の瞬きを繰り返すようにしてもよいことはいうまでもない。

【0077】図11および図12は、図9および図10に示したタイムチャートを上述した人物画像表示装置に実行するためのフロー図である。なお、図11、図12は、それらが合わさって一つの図を構成するようになっている。また、このフローの実行に先立ち、図13に示すロ変数テーブル（a）、目変数テーブル（b）を作成しておき、画像データベース746に格納しておく。

【0078】以下、ステップ順に説明する。

【0079】ステップ1（S1）
まず、初期段階として、ロの変数ｍｏｕｔｈ＝0、目の状態変数ｅｙｅ＝0としておく。

【0080】ステップ2（S2）
ｍｏｕｔｈ＝0、ｅｙｅ＝0にそれぞれ対応する口画像および目画像を出力する。この場合、それぞれの画像は閉じた口および閉じた目となっている。

【0081】ステップ3（S3）
入力音声を取り入れ、それが有音であるか無音であるかを判定する。有音である場合はステップ4へいき、また、無音である場合はステップ14へいく。

【0082】ステップ4（S4）
ステップ3で判定した有音が2度目以上の有音が初めての有音であるかを判定する。初めての有音である場合はステップ5へいき、また、2度目以上の有音である場合はステップ6へいく。

【0083】ステップ5（S5）
変数tに乱数表（0，T_J、$2T_J$、……、T_M）から選

11

ばれた数値を代入する。瞬きの開始時間をタンダム化させるためである。

【0084】ステップ6（S6）
変数tをT_Eで割り算を実行し割り切れるか否かを判定する。瞬きをT_E毎に行わさせるためである。割り切れる場合はステップ7へいき、割り切れない場合はステップ8へいく。

【0085】ステップ7（S7）
ｅｙｅ＝1に相当する目の画像を出力する。そして、変数tをT_Jだけ増加させる。

【0086】ステップ8（S8）
ｅｙｅ＝0に相当する目の画像を出力する。そして、変数tをT_Jだけ増加させる。

【0087】ステップ9（S9）
ロの大きさが増加中か否かを判定する。増加中の場合はステップ10へいき、そうでない場合はステップ12へいく。

【0088】ステップ10（S10）
ロの大きさが最大値に達したか否かを判定する。達した場合にはステップ13へいき、達しない場合にはステップ11へいく。

【0089】ステップ11（S11）
ロの変数に1を加算し、その加算値に対応する口画像を出力する。その後、ステップ3へいく。ステップ10からステップ11までの動作は、ロの大きさが増大しているのであれば、その最大値までロの大きさを増大させるためである。

【0090】ステップ12（S12）
ロの大きさが最小値に達したか否かを判定する。達した場合にはステップ11へいき、達しない場合にはステップ13へいく。

【0091】ステップ13（S13）
ロの変数に1を減算し、その減算値に対応する口画像を出力する。その後、ステップ3へいく。ステップ12からステップ13までの動作は、ロの大きさが減少しているのであれば、その最小値までロの大きさを減少させるためである。

【0092】ステップ14（S14）
2度目以上の無音が初めての無音か否か判定する。初めての無音の場合はステップ15へいき、2度目以上の無音の場合はステップ16へいく。

【0093】ステップ15（S15）
変数tに乱数表（T_J、$2T_J$、……、T_E'）から選ばれた数値を代入する。瞬きの開始時間をタンダム化させるためである。

【0094】ステップ16（S16）
変数tをT_E'で割り算を実行し割り切れるか否かを判定する。瞬きをT_E'毎に行わさせるためである。割り切れる場合はステップ17へいき、割り切れない場合はステップ18へいく。

12

【0095】ステップ17（S17）
ｅｙｅ＝1、0、1となる目の画像を順次出力する。それらの出力はT_J間隔となり、その後において、tに3T_Jを加算する。

【0096】ステップ18（S18）
ｅｙｅ＝0の目の画像を出力する。その後、tにT_Jを加算する。

【0097】ステップ19（S19）
無音期間中はｍｏｕｔｈ＝0としてロを閉じておく。そして、ステップ3へいく。

【0098】実施例3，図14は、上述した音声駆動を用いた人物画像表示装置をネットワークを介して他の音声駆動を用いた人物画像表示装置とインタラクティブに対話を行う場合の実施例を示したブロック図である。

【0099】同図において、図1と同符号のものは同一の機能を有する。

【0100】図1と異なる構成は、一方の人物画像表示装置側の対話者が発する音声は音声入力部71を介してデータ送信部7431に送られ、さらにネットワーク75を介して他方の人物画像表示装置側のデータ受信部7432に送られるようになっている。

【0101】そして、他方の人物画像表示装置の制御部では、該データ受信部7432によって受けた音声信号から、図1の場合と同じように、その言葉の意味を判断し、対話処理部742によって対話者の意図を汲み取り、かつ、それに対する返答を判断するようになっている。

【0102】音声解析部744では、データ受信部7432からの音声に対してたとえばＴｉの時間間隔ごとに該音声の変化を取りだし、その強さの平均値の算出、あるいはその他の必要な演算処理を行うようになっている。

【0103】そして、この音声解析部744で得られた音声の時間毎に区切られた各平均値等の時系列的な各情報は画像管理部745に出力され、この画像管理部745では、前記それぞれの情報に応じて画像データベース746から予め定められた適当な画像が順次選択され、この画像は画像出力部73に出力されて順次表示されるようになっている。

【0104】この画像は、前記対話者の音声に対して返答をする人物の画像（一人の場合もあるし、複数の場合もある）が主となり、その人物のロあるいはそぶり等は、前記音声出力部72から発せられる音声に追従して（同じタイミングで）動くようになっている。

【0105】この場合において、ネットワークを介して接続されるそれぞれの人物画像表示装置における音声解析部744、画像管理部745、および画像データベース746は、実施例1にて示した基本アルゴリズムおよびそれをさらに拡張させた拡張アルゴリズムに基づいて機能することはいうまでもない。

13

【０１０６】また、このような構成は、たとえばインターネットフォンのように、パソコンで通信しながらお互いの音声を用いて会話するようにすることもできる。この場合、音声だけでなく、画像データベース７４６に登録してある顔画像（たとえば話しをする相手の顔画像）を自分のパソコンの画面上に表示する。データ受信部７４３２から送られてきた音声の強弱などの情報を音声解析部７４４で求め、その結果を受けて画像データベース７４６から適当な画像が順次選択され、画像出力部７３に出力される。

【０１０７】以上、上述した各実施例によれば、人物の表情において、たとえばその口の場合に、話し始めで口が開き話し、さらに終わりで口が閉じるようになり、また声の大きさ等で分類した数に応じてそれらの間で口の開きの大きさを変化させることができることから、ほとんどの対話者が不自然さを感じることなく観察することができるようになる。

【０１０８】したがって、特に、複雑な手法を用いる必要がないことから、専門の技術あるいは装置を必要とすることなくほぼ自然的に対話をしている状態を疑似的に実現できるようになる。

【０１０９】なお、上述した各実施例では、いずれも基本アルゴリズムおよび拡張アルゴリズムにしたがって動作するように説明したものであるが、これに限定されることはなく、それらの一方にしたがって動作できるようにしてもよいことはいうまでもない。

【０１１０】

【発明の効果】以上説明したことから明らかなように、本発明による音声駆動を用いた人物画像表示方法および装置によれば、専門の技術あるいは装置を必要としなくてもほぼ自然的に対話をしている状態を疑似的に実現できるようになる。

【図面の簡単な説明】

【図１】本発明による音声駆動を用いた人物画像表示装置の一実施例を示すブロック図である。

14

【図２】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図３】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図４】本発明の基本アルゴリズムの一実施例を示す説明図である。

【図５】本発明の拡張アルゴリズムの一実施例を示す説明図である。

【図６】本発明の拡張アルゴリズムの一実施例を示す説明図である。

【図７】本発明の拡張アルゴリズムの一実施例を示す説明図である。

【図８】本発明の一実施例におけるアルゴリズムを実験的に試みた際の画像を示す図である。

【図９】本発明の拡張アルゴリズムのさらに拡張された実施例を示すタイムチャートで、図１０とで一つの図を構成している。

【図１０】本発明の拡張アルゴリズムのさらに拡張された実施例を示すタイムチャートで、図９とで一つの図を構成している。

【図１１】図９および図１０に示すタイムチャートをフロー図で、図１２とで一つの図を構成している。

【図１２】図９および図１０に示すタイムチャートをフロー図で、図１１とで一つの図を構成している。

【図１３】図１１および図１２に示すフロー図の動作に先立って予め装置に格納させておくデータを示す説明図である。

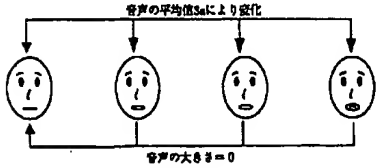
【図１４】本発明による音声駆動を用いた人物画像表示装置の他の実施例を示すブロック図である。

【符号の説明】

７１……音声入力部、７２……音声出力部、７３……画像出力部、７４……制御部、７４１……音声認識部、７４２……対話処理部、７４３……音声合成部、７４４……音声解析部、７４５……画像管理部、７４６……画像データベース。

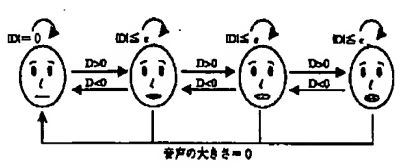
【図３】

図３



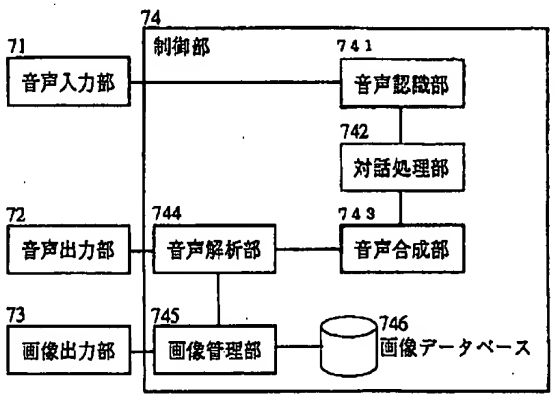
【図６】

図６



【図１】

図１



【図１３】

図１３

口変数と目変数テーブル

(a)

口変数の値と口画像

口変数（mouth）	口画像
0	
1	
2	
3	
・	・
・	・
N	

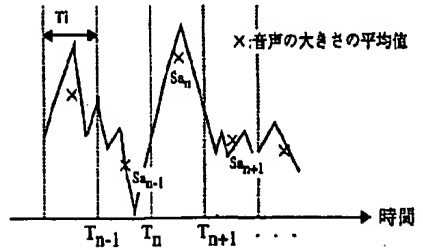
(b)

目の状態変数と目画像

目の状態変数（eye）	目画像
0	
1	

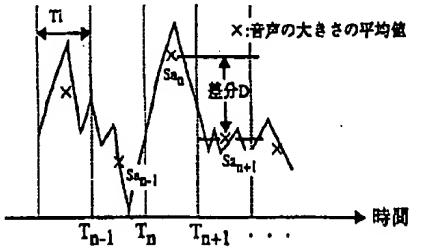
【図２】

図２



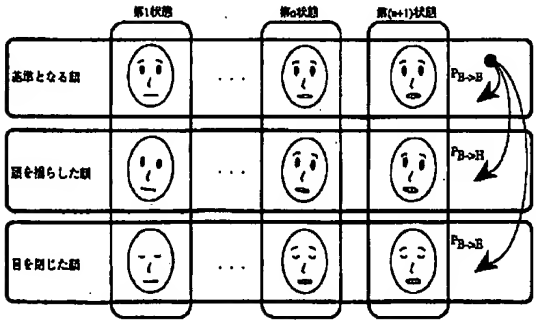
【図５】

図５



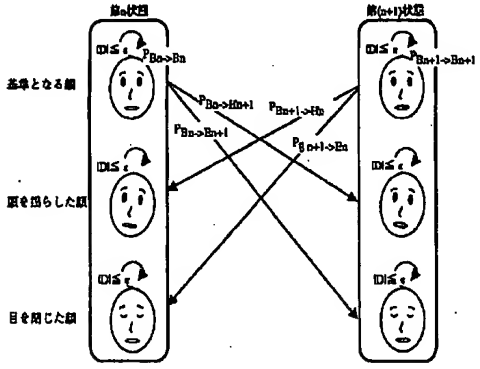
【図４】

図４



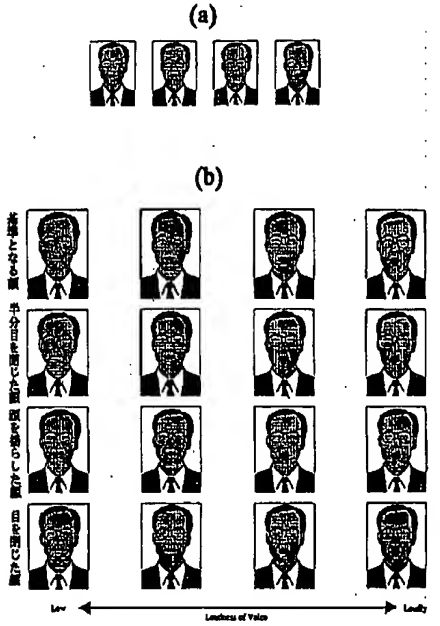
【図７】

図７



【図８】

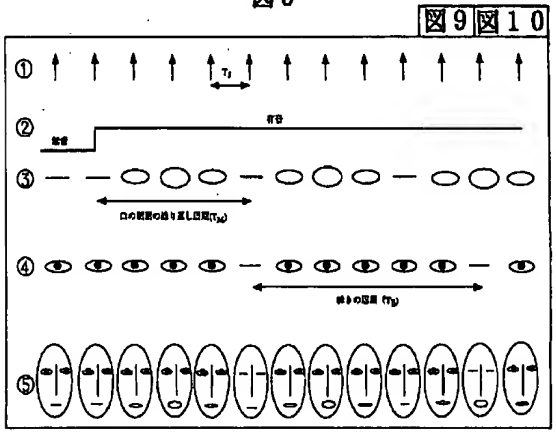
図８



【図９】

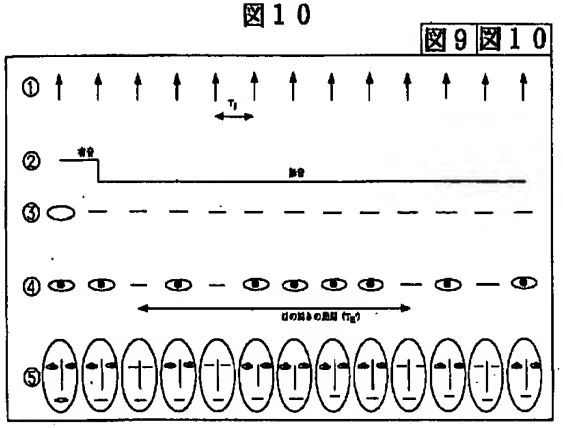
図９

図９ 図１０



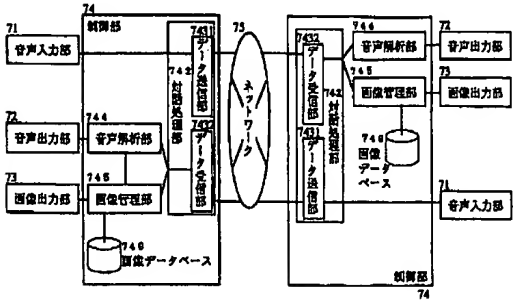
【図１０】

図１０



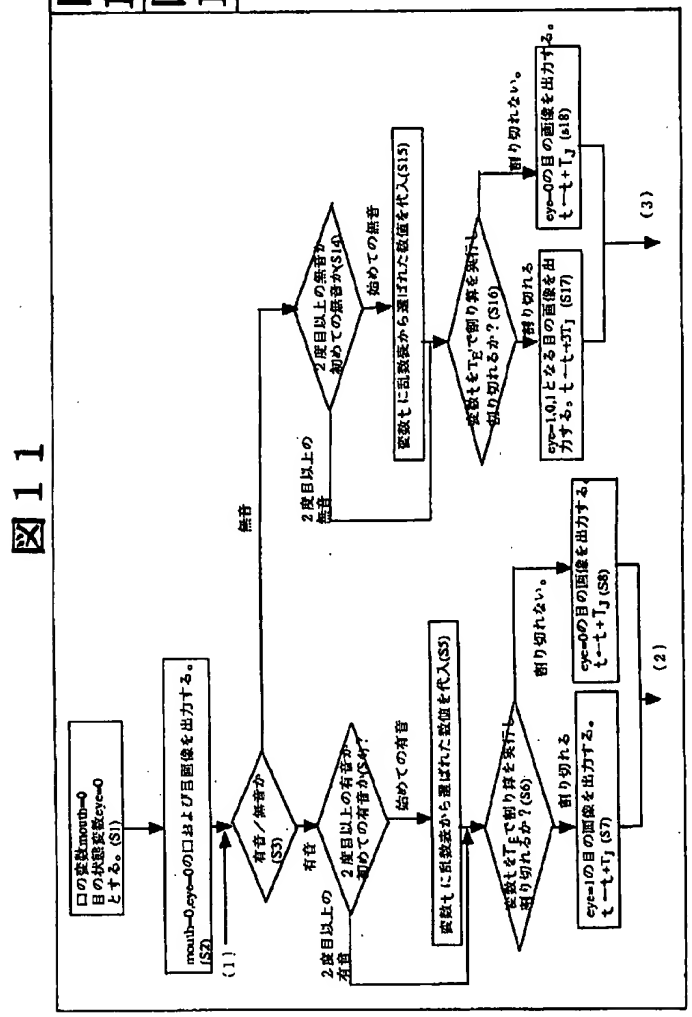
【図１４】

図１４

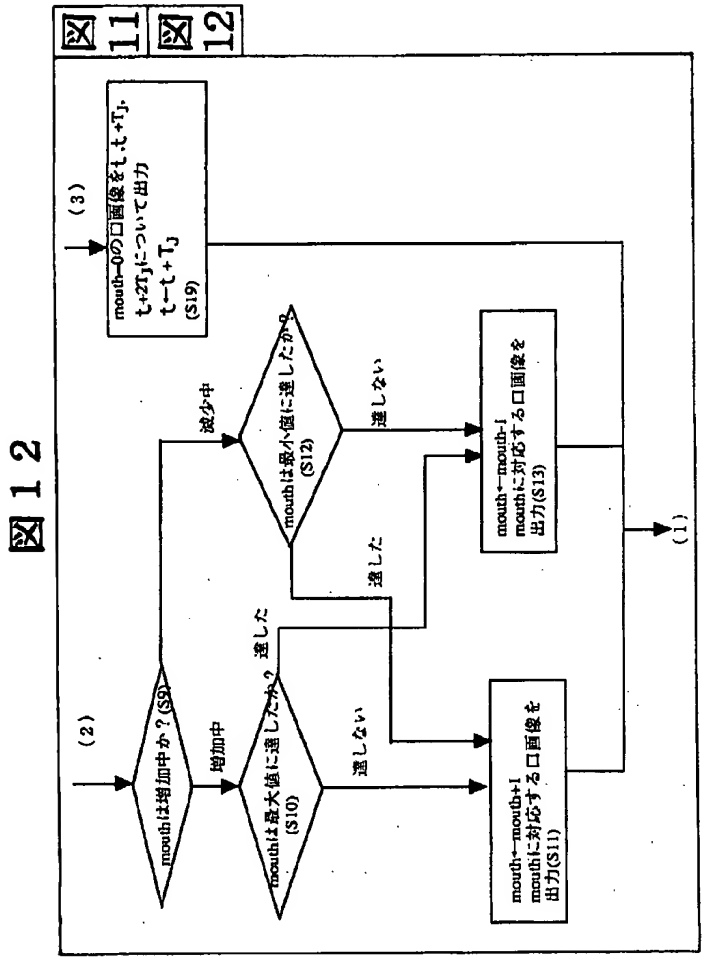


【図１１】

図１１ 図１２



【図１２】



フロントページの続き

(72)発明者　渡辺　信之
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内

(72)発明者　金山　英明
東京都新宿区西新宿三丁目19番２号　日本電信電話株式会社内